Kant's Theory of Freedom

Jonathan Bennett

(A commentary on Allen W. Wood, *Kant's Compatibilism*, in the same volume.)

Great knowledge, skill, and judgment have gone into Allen Wood's extraction from Kant's texts, and partial defence, of a certain theory of freedom (see preceding essay). I shall later mention one respect in which I am not sure he has got Kant right, but otherwise the interpretation is flawless. I shall argue, however, that although it is worthwhile to identify Kant's theory of freedom as Wood has helped us to do, the theory itself is worthless. I shall not list the reasons that Wood anticipates being brought against the theory. I do have those too, being unconvinced that the concepts of noumenon and of timeless agency are really intelligible. When Kant says of a noumenon that "nothing happens in it" and yet that it "of itself begins its effects in the sensible world" (B 569), he implies that there is a making-begin which is not a happening; and I cannot understand that as anything but a contradiction. Kant himself has trouble relating timeless choices to the temporal world. On the one hand, "at the point in time when I act, I am never free" (KPV 94g 98e); on the other, "In the moment when he utters the lie, the guilt is entirely his" (B 585). Never mind. For present purpose I concede noumena, timeless agency, non-Humean causation - *the lot*. With all of that granted, the theory is still worthless.

According to the theory, a free choice by my intelligible character causes me to have empirical character E. How can this be so, if there is also a deterministic causal explanation for my possession of E? How can a free choice cause this part of the natural causal chain without breaking the chain? Wood answers on Kant's behalf that my intelligible choice causes not only my possession of E but also a complete natural causal history for my possession of E. Kant didn't ever actually say this but Wood thinks that Kant's theory "must" be construed in this way. I'm not sure that it must, but in the meantime I shall assume that it is.

One significant fact about my character E is that I have beliefs about the Holocaust. These beliefs were partly caused by the Holocaust. Does Kant's theory make me responsible for what was done to the Jews of Europe when I was a child? Not necessarily, says Wood. In his version of Kant's theory, what I am morally responsible for is not the actual causes of my having character E but rather "those events which *must* belong to the actual course of things because I have the empirical character . . . that I do" (italics added). The actual Holocaust does not satisfy that condition if I could have had those beliefs in a world in which they were false and there was no Holocaust. And similarly for every event that has helped to shape me and over which we would ordinarily say I had no control: perhaps each and every one of them is inessential to my having character E.

I shall not discuss that as it stands. Kant is not out of trouble here unless Wood's defence works not merely for events but also for states of affairs. We have to consider the set of possible worlds where I have character E, and ask whether there are any remotely past states of affairs which obtain - that is, propositions that are true - at all of them but not at all deterministic worlds whatsoever. One might think, for example, that in 1929 oxygen exists not in all deterministic worlds but in all the ones where I am born in 1930 with character E; and so by Kant's theory I am morally responsible for the presence of oxygen in the universe in 1929. Wood's defence must be to suppose that neither that nor any other prenatal state of affairs is causally required for my

having character E: it was causally possible for me to burst onto the scene in 1930 *whatever* the scene was like before I arrived.

That makes the supposed choice of a causal history safe by making it vacuous; and Wood seems to intend it to do so. But it makes other things vacuous as well, draining all the content out of the notion of causal order. The thesis that every possible prenatal state of affairs is causally compatible with my having character E would be merely silly unless it was based on the general thesis that every possible sequence of states of affairs falls under some set of causal laws. But that thesis is not available to Kant: it would make nonsense of, at least, his Second Analogy and of his inference from determinedness to predictability.

Wood appeals to our ignorance of "how our timeless choices operate on the temporal world," but that does not help. The theory is that our free choices result in some present states of affairs, and also that choosing a certain state of affairs involves choosing a complete deterministic causal history for it. From those two bits of the theory, and a proper understanding of determinism, it follows that we freely choose states of affairs that antedate our births. This is a *proof* that these so-called timeless free choices are nothing like exercises of moral responsibility, and thus that Kant's theory about them is worthless.

The foregoing tells against a version of Kant's theory for which there is no direct textual evidence. Perhaps Kant himself would try in some less fatal way to reconcile my empirical character's being naturally caused with my freely choosing to have it. That may be what he is doing when he describes as "an effect of intelligible causality" not *this causal chain* but "*this empirical causality*" (B 572, italics added), suggesting that what 1 freely choose are not the causally interrelated items but rather the causal relation that links them. Wood quotes this passage, but I am not sure what he makes of it. He certainly does not remark that it might offer an alternative to his fatal solution to Kant's problem. I must admit that if it is an alternative solution I don't really understand it; still, it reminds us that Kant may have other things up his sleeve, so that we ought not to condemn his theory merely on the strength of Wood's rather creative version of it.

Here is a fresh proof that the theory is worthless. Never mind the natural causes of my having character E; let us simply consider what the theory says about my responsibility for my having E itself. Suppose that one fact about the kind of person 1 am is that I am insane. Wood quotes Kant as implying that in that case I cannot act autonomously, which must mean that my insanity is not a consequence of noumenal choice on my part (see preceding essay and VM 182). But Kant could not conceivably have grounds for saying that or, more generally, for supposing that the results of noumenal freedom come anywhere near to coinciding with the matters for which we regard ourselves as morally responsible. Even if we do not - as apparently Wood does not - hold Kant to his statement that "reason is present in all the actions of men at all times and under all circumstances" (B 584), Kant still has no basis for distributing reason through the actions of men in a manner acceptable to us, for example, for denying that noumenal freedom shines brightest in the daily doings of small babies and schizophrenics. It is essential to his theory that nobody could possibly have grounds for any claim about what range of empirical facts is attributable to noumenal freedom - apart of course from claims based on the requirements of consistency.

From the fact that Kant associates noumenal freedom with "reason", one might infer that someone who manifestly cannot reason lacks noumenal freedom; but that inference would be mistaken. Kant does tie his theory to a contrast between "sensuous impulses" (B 562) and "understanding and reason" (B 574-5), but the former include every motivating episode that has

natural causes, and the latter - understanding and reason - are described by Kant as faculties that "we distinguish . . . from all empirically conditioned powers." The line he is drawing, then, does not cut through the empirically given facts; rather, it has all the given facts (including empirical-world reasoning) on one side, and the dark noumenal theory (including otherworldly reasoning) on the other.

Wood doesn't dispute any of this. In face of it, he offers us the possibility that noumenal freedom has a scope that corresponds to what we ordinarily take to be the scope of moral responsibility. "It seems open to Kant," he says, "to suppose that [freely chosen events] correspond to those events for which we normally regard ourselves as morally responsible." This is backed by the observation that all Kant aims to establish is a possibility.

Wood is right to point out that sometimes Kant accepts the complete divorce of his theory from the world of human conduct as we experience it, and says that he wants to establish not even a real possibility but merely a lack of self-contradiction in our beliefs about freedom (B xxix, B 586). But Kant is not candid about how many other beliefs about freedom are equally vindicated by his theory. For example, in the famous discussion of the malicious lie (B 582-3), it is significant that he takes a *malicious* lie whose natural causes include "the viciousness of a natural disposition insensitive to shame," thus inclining us to agree that the lie's causes do not excuse the agent - "the guilt is entirely his." But Kant's theory allows us to pass that judgment not only in this case but also in one where the natural causes of the lie involve a profound psychopathology in someone who is not vicious and is greatly given to shame; and Kant's choice of example seems designed to help us to overlook that fact.

Even more striking are the places, not mentioned by Wood, where in the very act of declaring his theory's empirical emptiness Kant sneaks some content into it. For example, he says that because we don't know how much to attribute to noumenal freedom "the real morality of actions . . . remains entirely hidden from us" (B 579n); but Kant doesn't mean this as radically as he ought to, for he adds that therefore "no perfectly just judgments can be passed" on anyone's empirical character. The suggestion that we can at least approximate to justice is something to which Kant is not entitled. A second example: Kant says that someone's "intelligible character can never be immediately known" and that It must "be thought" (B574). So far, so good, but there is more - the intelligible character must "be *thought* in accordance with [*gemäss*, in agreement with, by the measure of] the empirical character," which implies that we know something about how the empirical relates in detail to the noumenal. A third example of Kant's giving in one phrase what he takes away in the next is his saying that a person's intelligible character "is completely unknown" and then adding "save in so far as the empirical [character] serves for its sensible sign" (B 574)·

I do not dispute that insofar as Kant had a single doctrine about noumenal freedom it was the one Wood attributes to him, namely, that our beliefs about freedom do not logically conflict with determinism; that determinism should obtain and yet those beliefs be true is logically possible. But I protest that this kind of possibility is not worth establishing, as may be seen from the fact that endless other sets of opinions are also shown by Kant's theory of freedom to be possible - for example, that only madmen and babies are morally responsible - and the theory provides no means for adjudicating amongst the sets.

Lewis White Beck even questions whether this is a theory about "freedom" in our ordinary sense of the word. What the theory provides, he says, is "not what is meant by freedom in any interesting sense, because it is indiscriminately universal," his point being that it "seems to justify

the concept of freedom, if anywhere, then everywhere."[1] Wood responds to this objection by setting aside Kant's "reason is always present" remark and offering a version of the theory which says nothing about the scope of noumenal freedom, thus allowing that freedom could be present in exactly the cases in which we intuitively think it is. That undercuts Beck's premise, but his conclusion still stands. As well as being silent about the scope of freedom, Kant's theory implies that we cannot have grounds for any specific opinion about what that scope is; and that fact debars it from being about 'freedom' in the ordinary sense just as would the theory's implying that most of our opinions about the scope of freedom are false.

Anyway, let us keep sight of the fact that although the Kantian theory says that our untutored opinions on freedom might be right, it offers no way in which that could be other than sheerly fortuitous - no suggestion about how the truth of those opinions might help to explain why we have them. In the absence of that, the possibility that they are true is of no interest.

So much for Kant's official central theory. Most of us have long thought it to be dead, and after Wood's restorative measures the corpse still refuses to stir. What remains to be considered is the thick detail of Kant's live thinking about freedom, and in conclusion I want to say a little about that. This is a matter on which Wood and I disagree sharply. I see myself as turning from what is dead in Kant's writings to what is alive; as setting aside Kant's map and getting out into the countryside with him; but when I did that in *Kant"s Dialectic* it convinced Wood - as he reported in his review of the book - that "Bennett is not really much interested in Kant's philosophy itself." Wood's use of the phrase "Kant's philosophy itself"expresses a view about what we should be looking to Kant for, a view I believe is wrong. His own opinion notwithstanding, Kant was bad at grand theory construction. Where he was superb was in the informal discussions surrounding the theories: his sensitivity and subtlety of response to conceptual pressures and tensions make those discussions wonderfully instructive, though these same qualities lead him into intricacies and inconsistencies. In ignoring the latter, Wood ignores Kant's greatest strengths as a philosopher. Turning his back on what doesn't fit the large theoretical structure, which he calls "Kant's philosophy itself," he is turning away from the life in Kant's text in order to preside over a corpse.

In *Kant"s Dialectic* I attended not only to Kant's official theory but also, more fully, to various hints and indications in his informal discussions. The following sketch outlines my strategy and contrasts it with Wood's approach in the preceding essay.

I agree with Kant: there is an apparent clash between a freedom thought and a determinism thought, and the two are reconcilable because of some difference in angle or level or standpoint. But I hold that there are two distinct clashes, and two reconciliations; and although Kant did not consciously notice this, the two show up very differently in his text, a fact that indicates his sensitivity to the difference between them.

One of the two conflicts concerns moral accountability, which surfaces in Kant's discussion when he speaks of blame and guilt and of what didn't happen but "ought to have" (B 562, B 578, B 582-3). Most of us find that our propensity to blame can be made to look unfair by its being brought up hard against the hypothesis of determinism - and yet somehow most of us think that we would sometimes blame people even if we believed determinism were true. I find Kant valuable as a pointer to the acuteness of this problem and to the unacceptability of the standard shallow kind of compatibilism. In *Kant"s Dialectic* I sketched a possible solution to this, based on

---

[1] Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago, 1960), p. 188.

P. F. Strawson's great essay, "Freedom and Resentment."[2] Here is the core of it. Start with the idea of my resentment of something you have done to me. Then consider a case in which you have done something I dislike, though not essentially because of how it affects me, and I hold it against you, adopting an attitude like resentment except that it lacks the essential reference to myself. This 'vicarious resentment', as it might be called, is blame. Strawson holds that all our praise- and blame-related responses to human conduct should be understood as developments from the more personal attitudes and feelings of gratitude and resentment. The nature of these personal 'reactive attitudes' and their role in our lives explain why they are inappropriate under some conditions, for example, why it is unsuitable to resent a baby's disturbance of one's sleep by its crying. That then lets us explain the various conditions under which praise and blame are inappropriate: if the conditions are not satisfied, what follows is not that *judgments* involved in praise and blame are false but rather that *feelings and attitudes* involved in praise and blame are inappropriate. These explanations have not the slightest tendency to imply that if determinism is true then all praise and blame is wrong, but Strawson puts us in a position to explain why determinism is often thought to be a threat to praise and blame. Here is how.[3] The personal reactive attitudes which are the home ground of praise and blame are, in large measure and for most people, in some sort of conflict with the objective attitude in which one seeks to gather the facts, to understand the situation, to discover the etiology of the behaviour so as to alter its chances of recurring. Gratitude for a gift, for instance, does not sit easily in the mind alongside an attitude of active inquiry into the gift's causal origins. Thus reactive attitudes are in conflict or tension with the frame of mind in which one so much as raises the question of an action's causal nature: it is not that determinism logically conflicts with blameworthiness, but rather that the raising of the question of determinism conflicts with the feelings and attitudes that go into blame and make it what it is. That also explains, less damagingly than is otherwise possible, why many people who think that if determinism is true then no one is blameworthy are also apt to think that if determinism is false then still no one is blameworthy.

This shares certain abstract features with Kant's theory of freedom: it is deep and systematic and does not accuse the incompatibilist of mere conceptual muddle; and it denies that a full treatment of blameworthiness can be given purely in terms of our perception of the given facts. The big difference between Strawson and Kant is that whereas Kant's theory ties blameworthiness to a thought of ungiven facts, Strawson's says that we must go outside all the facts and introduce a dimension of feeling.

It was Kant who taught me that there is a second prima facie conflict between freedom and determinism. Rather than blame for an action already performed by oneself or someone else, this conflict concerns practical deliberation by an agent wondering what to do. In a nutshell, deliberation involves viewing some questions about the future as radically open, while determinism seems to imply that they are all really closed. Kant presents this matter in a wonderful paragraph which says that "reason does not here follow the order of things as they present themselves in appearance, but frames for itself . . . an order of its own [and] presupposes that it can have causality in regard to all these actions" (B 575-6).

---

[2] Jonathan Bennett, *Kant"s Dialectic*, sections 66-7; P. F. Strawson, "Freedom and Resentment," in his *Freedom and Resentment and Other Essays* (London, 1974), pp. 1-25.

[3] I am relying partly on a development of Strawson's ideas in my "Accountability," in Z. van Straaten, ed., *Philosophical Subjects* (Oxford, 1980), pp. 18-47, especially pp. 25-8.

Because this is an essentially first-person problem, it is in the context of it that Kant links the concept of noumenon to how "man knows himself" (B 575) Also, because the problem is about future actions rather than past ones, it is here that Kant speaks of "imperatives" and of what "ought to be" (B 575-6) rather than of what "ought to have happened."

In *Kant"s Dialectic* I defended a treatment of this matter stemming from an insight of Ryle's.[4] Kant invites us to contrast (a) following the order of things as they present themselves with (b) framing for oneself an order of one's own and, as he sometimes says, acting under the idea of freedom. What *is* this contrast? What would it be to behave in that manner (a) from which Kant's theory of freedom is supposed to rescue us? The only clear, literal sense I can make of it is to suppose that (a) involves looking at one's future not as a deliberating and deciding agent but rather as a predicting self-observer who tries to work out what he will do by applying causal laws to his known present condition. Determinism threatens us with this by implying that there always is a sufficient basis for such a prediction, that is, that each of our actions was in principle susceptible of being soundly predicted in advance of its happening. We do not know enough actually to do this, of course, but we cannot be comfortable with the thought that our practical deliberations are a *pis aller,* that our status as deliberating agents is a pure product of ignorance. That is what Kant offers to rescue us from. He is saying that our deliberating stance is securely and deeply grounded, that it is not in danger - even in principle - of being swept aside by an inrush of knowledge of our structure and the laws that govern us.

If that is not what is involved in Kant's contrast between (a) and (b), then I do not know what is. I can find no clear alternative to it in his text or in the secondary literature.

If my interpretation is right, then Kant is right - or may very well be. As Ryle pointed out decades ago, the existence of facts about me that would warrant confident predictions about how I will act does not imply that it is possible, even in principle, that I should make such predictions. The facts on which a given prediction would have to be based might not be facts if I were thinking the prediction; and so there may be narrow limits on how much self-prediction it is in principle possible for me to do, whether or not determinism is true and however much knowledge I acquire.

Thus, we can unapologetically regard ourselves *fully* entitled to approach our futures in a deliberating rather than a predicting manner. The thought "I can do this only because I am ignorant" can be dismissed as a scare story that is very likely false even if determinism is true. So a certain peculiarity in the notion of self-prediction serves as a barrier which - without invoking noumenalist metaphysics - prevents the determinist thought from conflicting with the deliberating agent's belief that some options are still 'open' in the sense that the decision amongst them must be approached in the deliberator's and not the predictor's way.

This again has significant points in common with Kant's theory. In particular, it performs a reconciliation by bringing in not only the viewpoint of the observer of the given facts but also another viewpoint; with the difference that I make it the viewpoint of the agent, not that of the thinker of the ungiven facts.

In my handling of these two problems - accountability and agency - Kant's contrast between the observer's point of view and the noumenal point of view is replaced by a pair of contrasts: one

---

[4] Bennett, *Kant"s Dialectic*, sections 68-9; Gilbert Ryle, *The Concept of Mind* (London, 1949), p. 197.

between the observer and the emotional responder, the other between the observer and the agent.[5] You may protest that this is mere word-play, and irrelevant to what is really going on in Kant's pages. I think otherwise: 1 think that by reading Kant dispassionately and thinking hard about what he says - including many of the little twists and turns of phrase - I have learned some philosophy that is not contained in the frozen object Wood calls "Kant's philosophy itself." I submit that this approach is better: it contributes more to learning the truth, avails itself more fully of the Kantian texts, and does more honor to the genius of their author.

---

[5] In an unpublished paper entitled "Freedom and Objectivity,"' which is to he part of a forthcoming book, Thomas Nagel also treats freedom as generating two problems, one about accountability and one about agency. and he treats Strawson as crucial to the former, and Ryle's self-prediction point as crucial to the latter. Nagel's insightful treatment of self-prediction finds in it a moral significance which I did not see there,